TATA Binding Proteins Can Recognize Nontraditional DNA Sequences

Sunmin Ahn,[†△] Chia-Ling Huang,[‡] Emre Ozkumur,[§] Xirui Zhang,[†] Jyothsna Chinnala,[†] Ayca Yalcin,[§] Sabita Bandyopadhyay,[¶] Shelley J. Russek,[¶] M. Selim Ünlü,[†§] Charles DeLisi,[†‡||*] and Rostem J. Irani[||△]
[†]Department of Biomedical Engineering, [‡]Bioinformatics Program, and [§]Department of Electrical and Computer Engineering, Boston University, Boston, Massachusetts; [¶]Department of Pharmacology and Experimental Therapeutics, Boston University School of Medicine, Boston, Massachusetts; and [||]Center for Advanced Genomic Technologies, Boston, Massachusetts

ABSTRACT We demonstrate an accurate, quantitative, and label-free optical technology for high-throughput studies of receptor-ligand interactions, and apply it to TATA binding protein (TBP) interactions with oligonucleotides. We present a simple method to prepare single-stranded and double-stranded DNA microarrays with comparable surface density, ensuring an accurate comparison of TBP activity with both types of DNA. In particular, we find that TBP binds tightly to single-stranded DNA, especially to stretches of polythymine (poly-T), as well as to the traditional TATA box. We further investigate the correlation of TBP activity with various lengths of DNA and find that the number of TBPs bound to DNA increases >7-fold as the oligomer length increases from 9 to 40. Finally, we perform a full human genome analysis and discover that 35.5% of human promoters have poly-T stretches. In summary, we report, for the first time to our knowledge, the activity of TBP with poly-T stretches by presenting an elegant stepwise analysis of multiple techniques: discovery by a novel quantitative detection of microarrays, confirmation by a traditional gel electrophoresis, and a full genome prediction with computational analyses.

INTRODUCTION

Transcription initiation in eukaryotes is a complex process that involves sequence-specific interactions between multiple transcription factors (TFs) and the core promoter leading to the assembly of the preinitiation complex (PIC). Most eukaryotic genes whose core promoters are located 40–50 basepairs (bp) upstream and/or downstream from the transcription start site (TSS) are transcribed by RNA polymerase II (1). Multiple core promoter elements have been identified, including the TATA box, INR (the initiator), BRE (the TFIIB recognition element), DPE (the downstream promoter element), and CpG islands (2). Among these, the TATA box, which has the consensus sequence TATA(A/T)A(A/T)(G/A) (1), has been the most extensively studied. According to the textbook model of transcription initiation, the PIC assembly begins with the TATA-binding protein (TBP) directly binding to the TATA box 25–35 bp upstream of the TSS, followed by the recruitment of general/basal TFs and RNA polymerase II (1,3). However, a genome-scale computational analysis revealed that the majority of human genes (~76%) lack the TATA box in the core promoter (4). Recent studies suggested that the mechanisms of transcription initiation in eukaryotes are more diverse than the textbook model (2,3,5), and that the PIC can assemble with only a subset of TBP-associated factors (TAFs) for TATA-less promoters (6).

Is TBP to be disregarded in transcription initiation for genes with TATA-less promoters? There is some evidence that TATA box binding proteins can also recognize sequences lacking the TATA motif, albeit with substantially reduced affinity (7). Furthermore, the emergence of broad-peak TSS-containing promoters, where the TSS is distributed over a region of 100 bp (8), suggests PIC sliding or multiple PIC formation (5). As we try to uncover the diverse mechanisms of eukaryotic transcription initiation, particularly for vertebrates, it is reasonable to question whether alternative TBP binding motifs exist. Alternative binding motifs could be present in the core promoter or in broad-peak TSS-containing promoters that extend beyond the traditional core, which would lead to PIC assembly for TATA-less promoters.

A systematic search for alternative TBP binding sequences would be difficult to carry out with traditional with gel-based assays, because they are too labor-intensive for surveying numerous DNA and protein interactions. Alternatively, several solid-phase, array-based methods that allow multiplexed detection have been developed to study DNA-TF interactions (9–13). However, they present limitations due to the nature of solid-phase interactions. Here we apply a recently developed and validated solid-phase, high-throughput DNA-protein complex detection platform that 1), mimics solution-phase interactions; 2), is quantitative; and 3), is independent of the conformation of the molecules on the surface. In particular, we employ a label-free detection platform, interferometric reflectance imaging sensor (IRIS), to study TBP interaction with multiple types of DNA.

IRIS satisfies all of the aforementioned requirements. First, it uses a three-dimensional (3D) polymeric network (14) that swells up to 20 nm from the substrate surface upon hydration (15). This swelling of the polymer provides a solution-like microenvironment, thereby preserving

Submitted April 10, 2012, and accepted for publication August 14, 2012.

[△]Sunmin Ahn and Rostem J. Irani contributed equally to this work.

*Correspondence: delisi@bu.edu

Editor: David Millar.

0006-3495/12/10/1510/8 $2.00

http://dx.doi.org/10.1016/j.bpj.2012.08.030

DNA-TF affinity on the surface. Second, IRIS detection is based on optical path difference (OPD) measurements caused by surface accumulation of biomolecules, providing quantitative results. Protein-protein (16) and DNA-DNA interactions (17) were studied previously, and the conversion of OPD to mass density on the surface was reported (18). Finally, IRIS is independent of any conformational changes of the molecules on the surface. Double-stranded DNA (dsDNA) often undergoes conformational changes when interacting with proteins (19–21), and many label-free detection techniques based on evanescent optical fields are sensitive to surface conformation. In addition, protein binding to single-stranded DNA (ssDNA) can induce different conformational changes compared with dsDNA binding, which can further complicate quantification.

We use IRIS technology to investigate the interaction between TBP and ssDNA and dsDNA arrays that have similar probe density. We show that TBP binds specifically and tightly to T-rich oligonucleotides (polythymine (poly-T) stretches) and to single-stranded TATA (ssTATA) box sequences, as well as to traditional double-stranded sequences with the TATA box motif. The label-free results are confirmed with a traditional gel mobility shift assay, and the relative affinities of TBP are found by competition assays and IRIS. In addition, we present full human genome analyses of various locations of poly-T stretches and TATA box motifs within the extended core promoter, and discuss their implications for transcription initiation.

MATERIALS AND METHODS

IRIS detection principle

An outline of the detection principle (16,22) is shown in Fig. 1. Briefly, by using optical interferometry, we measure protein accumulation on the sensor surface by determining the optical path difference (OPD). We convert the measured OPD into mass density on the surface by using simple conversion factors (1 nm = 0.8 ng/mm^2 for ssDNA, and 1.2 ng/mm^2 for the additional protein layer) determined in previous work (18).

DNA microarray preparation

Silicon wafers with 500 nm of thermally grown oxide were purchased from Silicon Valley Microelectronics (Santa Clara, CA). The wafers were diced into approximately 15 mm × 15 mm chips and cleaned by sonicating in acetone for 5 min (three times), rinsing in methanol, then rinsing in deionized water. All chips were dried with nitrogen gas and cleaned with O_2 plasma asher. Copoly (N,N-dimethylacrylamide (DMA)-acryloyloxysuccinimide (NAS)-3(trimethoxysilyl)-propylmethacrylate (MAPS)), which will be referred to as copoly (DMA-NAS-MAPS), was synthesized and coated on the clean silicon chips as described elsewhere (23). All DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA) and purified by high-performance liquid chromatography to >80% purity. For the purpose of immobilization, some oligonucleotides had amine modification on their 5′ end. We investigated 15 different sequences with 10–20 replicate spots to demonstrate IRIS as a high-throughput platform to study TF. Each microarray consisted of four replicate arrays amounting to >600 DNA spots per chip. The sequences of the oligonucleotides relevant to this study were designed to result in minimal secondary structures (Table 1). All oligonucleotides were hybridized with complementary probes without the amine tag in 150 mM sodium phosphate buffer, pH 8.5. To ensure that all amine-modified oligonucleotides were in a duplex form, the concentration of the complementary probes was in 5% excess during hybridization. The hybridized dsDNA was spotted onto the functionalized surface by BioOdyssey Caligrapher MiniArrayer (Bio-Rad, Hercules, CA) at a final concentration of 20 μM in 150 mM sodium phosphate buffer, pH 8.5. Any excess oligonucleotide that did not have amine modification was washed off the array (24). The thermodynamic properties of the oligonucleotides were calculated with OligoAnalyzer (Integrated DNA Technologies). The free energies of all possible hairpin structures were found to be > −1.63 kcal/mol, which is significantly greater than that of the duplex (ΔG = −50.7 kcal/mol). Under these conditions, the intramolecular structure is inconsequential.

The spotted arrays were kept in a humid environment with 65% humidity overnight for immobilization reaction to complete. All chemical reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless noted

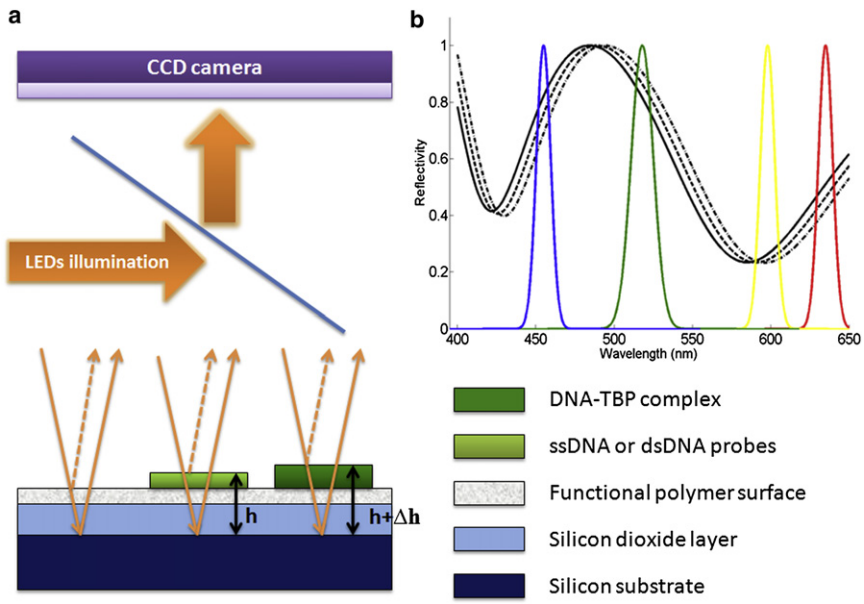


FIGURE 1 Description of label-free detection of DNA-protein interactions using IRIS. (*a*) Four LEDs with peak emission at 455 nm, 518 nm, 598 nm, and 635 nm are used to illuminate the biosensor surface. Light is reflected at both the air-SiO_2 interface and the Si-SiO_2 interface. Reflected light from the two optical paths interferes, and the intensity over the surface is imaged with a CCD camera. A functional polymeric surface is used to specifically immobilize DNA on the surface in a microarray format. Accumulation of biomolecules on the surface changes the optical path, causing a phase shift in the reflected intensity. (*b*) The reflected intensity of each pixel is normalized and fitted with an algorithm at four wavelengths. The difference in optical path length is mapped into a height change and is illustrated as a shift of the intensity curve.

TABLE 1 Oligonucleotide sequences used in the study, with the consensus sequence underlined

Name	Sequence
ssTATA	5′ – NH_2 GACCTCGGTATAAAAGGGCGCTGG – 3′
25-mer T	5′ – NH_2 TTTTTTTTTTTTTTTTTTTTTTTTT – 3′
25-mer A	5′ – NH_2 AAAAAAAAAAAAAAAAAAAAAAAAA – 3′
dsTATA	5′ – NH_2 GACCTCGGTATAAAAGGGCGCTGG – 3′
	3′ - CTGGAGCCATATTTTCCCGCGACG – 5′
ds25-mer dA/dT	5′ – NH_2 AAAAAAAAAAAAAAAAAAAAAAAAA – 3′
	3′ – TTTTTTTTTTTTTTTTTTTTTTTTT – 5′

otherwise. The DNA arrays were washed four times in 2X saline sodium citrate (SSC) buffer for 10 min each, and then 0.2X SSC for 1 min on a shaker, followed by two final rinses in 0.1X SSC. A number of the DNA arrays were washed three times in deionized water for 3 min to denature the dsDNA and make ssDNA arrays, as it was previously shown that deionized water can completely denature the double helix at room temperature (17). The microarrays were imaged with IRIS to determine the immobilization density of the spotted DNA probes.

TBP binding

Human TBP was purchased from Proteinone (Rockville, MD). The DNA microarrays were imaged with IRIS to acquire the surface density of the oligonucleotides before the binding experiments. The microarrays were treated with TBP for 2 h on a rotating shaker in a binding buffer at pH 7.0 that consisted of 10 mM HEPES, 70 mM KCl, 10 mM $MgCl_2$, 1 mM EDTA, and 0.9 M trehalose. To avoid aggregation of TBP, we used trehalose as a deaggregator (25). After TBP binding, the microarrays were washed in the same binding buffer without TBP 5 times for 3 min, followed by a final rinse in deionized water. Finally, the microarrays were imaged with IRIS to measure the OPD upon TBP binding.

Electrophoretic mobility shift assays

Purified TBP was used for all gel experiments. TBP was premixed on ice to a final volume of 20 μL containing the following components: 10 mM HEPES, 70 mM KCl, 5 mM $MgCl_2$, 1 mM EDTA, 0.1% Triton X100, 0.9 M trehalose, and 5% glycerol. The above reaction mixtures were loaded onto a 5% (60:1) polyacrylamide gel containing 5 mM $MgCl_2$, 5% glycerol, 2.5 mM dithiothreitol (DTT) in 0.5X Tris-borate-EDTA (TBE) buffer. The running buffer was 0.5X TBE. Xylene and bromophenol blue dyes were added to DNA lanes only, because they are known to interfere with DNA protein binding. All ssDNA sequences were purchased from Integrated DNA Technologies (Coralville, IA) without any modifications. The oligonucleotides were ^{32}P labeled using T4 polynucleotide kinase (PNK) and hybridized with unlabeled complement to form radiolabeled dsDNA. The amount of DNA used in one lane was 10,000 cpm or 1–5 ng, and TBP was used in excess 0.21 μg. Electrophoresis was conducted at 140 V (constant voltage) for 3.5 h at 4°C. The gels were vacuum dried for 1 h at 80°C on blotting paper. The gels were then exposed by conventional autoradiography, and 10X unlabeled competitor DNA was used for competition assays from radiolabeled DNA (see Fig. S1 in the Supporting Material).

Computational identification of TATA box and poly-T stretches in human promoter regions

We parse the extended core promoters for each gene in the human genome into four classes: 1), those with the TATA motif and without poly-T; 2), those having both the TATA motif and poly-T; 3), those with poly-T and no TATA motif; and 4), those with neither poly-T or the TATA motif. Human promoter regions from −2000 to 0 relative to the TSS (+1) of all human genes were downloaded from the UCSC genome browser (http://hgdownload.cse.ucsc.edu/downloads.html; hg19, GRCh37). A total of 17,920 genes were left after redundant entries were removed. We searched for poly-T stretches with at least nine consecutive thymines in the promoter regions from −2000 to 0. We searched TATA-like element in the core promoter regions (−50 to 0) to find correlations between poly-T stretches and the textbook model of the TATA box by using position-specific scoring matrices (PSSM) of the TATA box collected from the TRANSFAC database (26). We calculated score S for each position along the promoter region to evaluate whether a gene had a TATA-like element (27):

$$S = \sum_{i=1}^{L} \log\left(\frac{e_i(x_i)}{q_{xi}}\right)$$

where L is the length of the sequence being scored (the number of columns in the PSSM), i labels the position of a nucleotide in the sequence, $e_i(x_i)$ is the probability of observing nucleotide x in position i, and q_{xi} is the background probability of observing nucleotide x. S represents the log odds ratio of the scored sequence being derived from the TATA box motif versus being derived from the background sequence. A positive score S indicates that the scored sequence is more likely to be derived from the TATA box motif. We scored all overlapping sequences of length L within 50 bases of the TSS. With a permissive threshold ($S > 0$), 5486 genes with TATA-like elements in their core promoter regions were kept.

RESULTS AND DISCUSSION

TBP binds to ssTATA box, dsTATA box, and poly-T stretches

All DNA microarrays contained 16 replicate spots for each oligonucleotide. The arrays were imaged with IRIS before and after binding with TBP, as shown in Fig. 2. Four dsDNA arrays and four ssDNA arrays were prepared, and they were incubated with 5 μg/mL of TBP. The average measured heights of all types of DNA probes across the four microarrays are plotted in Fig. 3 *a*. The height increase of the probes

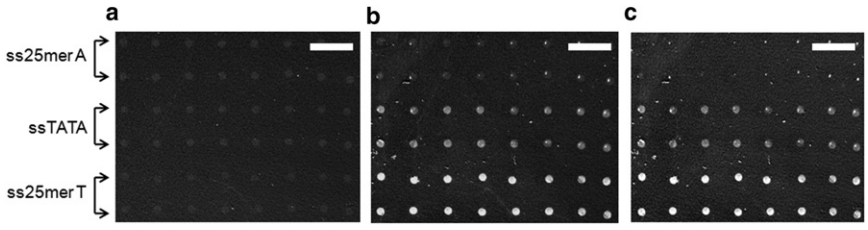
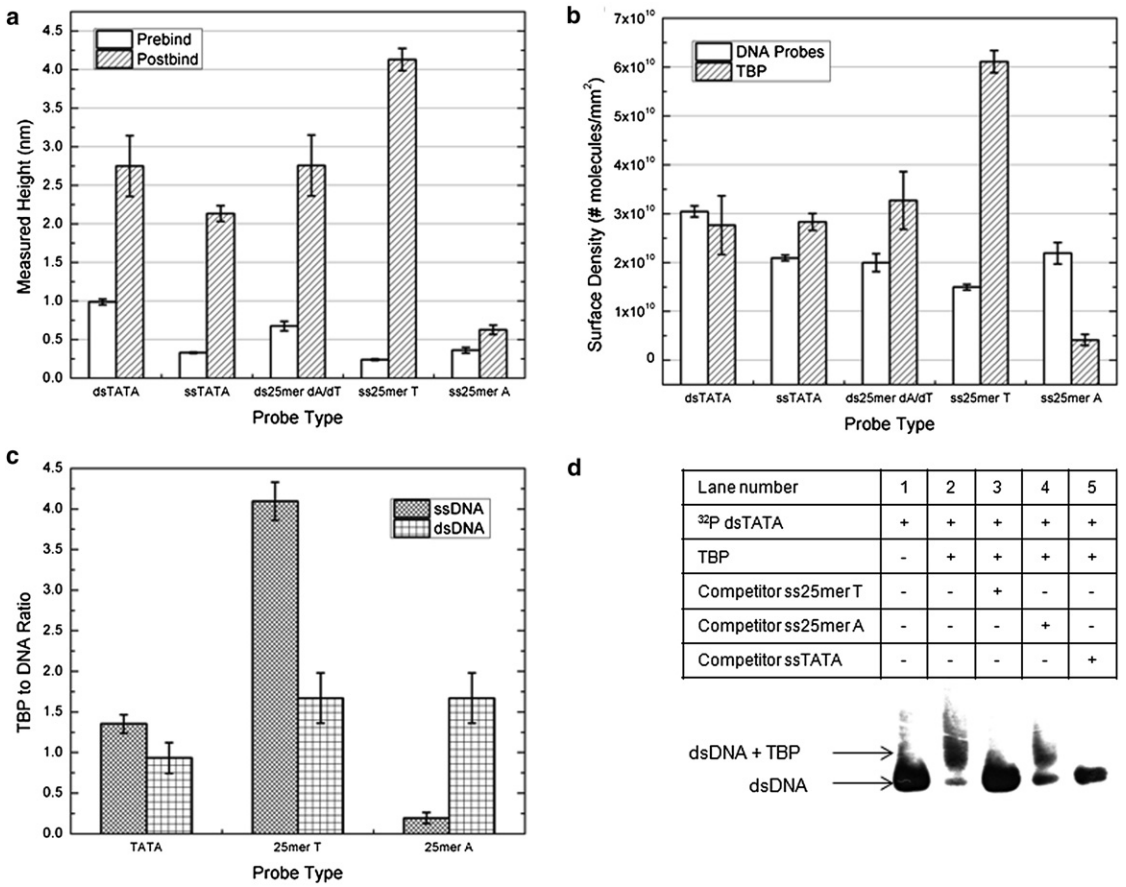


FIGURE 2 IRIS images of an ssDNA-TBP binding experiment. (*a*) Prebinding image of ssDNA microarray. Spot sizes are ~100 μm. (*b*) Postbinding image of ssDNA microarray after incubation with TBP at 5 μg/mL for 2 h. (*c*) Differential image of post- and prebinding images (postbinding image − prebinding image). Spot height compared to its local background is later quantified into biomolecule surface density. Scale bar, 500 μm.



Lane number	1	2	3	4	5
^{32}P dsTATA	+	+	+	+	+
TBP	-	+	+	+	+
Competitor ss25mer T	-	-	+	-	-
Competitor ss25mer A	-	-	-	+	-
Competitor ssTATA	-	-	-	-	+

FIGURE 3 (*a–c*) TBP interactions with ssDNA and dsDNA arrays. (*a*) Average height of the DNA spots before and after binding with TBP. The error bars represent ± 1 standard deviation of the average height of all the spots ($n = 16$) among four microarrays. (*b*) Calculated surface density of the molecules before and after binding with TBP. DNA probe density was calculated with the initial spot heights of the microarrays. The height increase upon TBP binding was used to calculate the surface density of TBP. (*c*) Ratio of the number of TBPs to DNA detected by IRIS. The ratio was found by dividing the surface density of the bound TBP by the surface density of the DNA probes. The ratio of TBP to ds25-mer dA/dT is plotted on both 25-mer T and 25-mer A, because 25-mer A and 25-mer T are perfect complements to each other. (*d*) IRIS results are confirmed by electrophoresis. Lane 2 shows a shift caused by the TBP-DNA complex. Competitions with ss25-mer T, ss25-mer A, and ssTATA in lanes 3–5 indicate that TBP binds to ssDNA in a sequence-specific manner.

after the addition of TBP shows that TBP binds to both ssDNA and dsDNA with the TATA motif. In addition, we observe a high number of TBPs binding to ss25-mer T, and very few binding to ss25-mer A. The measured optical thickness of the biolayers was converted to the average mass density of the biomolecules on the surface, which in turn was converted to molecular surface density using the appropriate molecular weights of each molecule (Fig. 3 *b*). The mass density of the dsDNA was normalized to account for the refractive index change between ssDNA and dsDNA in the same fashion as in our previous work (17).

Fig. 3 *c* shows the ratio of the number of TBPs to DNA detected on the surface. An ~1:1 ratio of TBP to DNA is found on both dsDNA and ssDNA with one TATA motif. Of interest, the ratio is very large for ss25-mer T (~4 TBP to DNA), indicating that multiple TBPs are bound to a single ss25-mer T probe. The presence of 0.9 M trehalose in the binding buffer reduces the possibility of TBP aggregation. We observe very little binding to ss25-mer A, which confirms that TBP does not bind to ssDNA indiscriminately and instead has a marked preference for T-rich sequences. The preference for T-rich sequence is also displayed by a higher ratio of TBP to DNA on ds25-mer dA/dT compared with that of dsDNA with one TATA motif. The IRIS results were confirmed by an electrophoretic mobility shift assay (EMSA), as shown in Fig. 3 *d*. The shift caused by the dsTATA-TBP complex is lost when the reaction is competed with cold ss25-mer T and ssTATA, whereas the shift is retained with cold ss25-mer A. TBP interactions with ssTATA, dsTATA, and ss25-mer T are confirmed to have high affinity compared with ss25-mer A.

For solid-phase methods, one must consider the surface density of the probes. Oligonucleotide surface density has been known to affect the kinetics as well as the thermodynamics of hybridization on the surface (28,29), and investigators have attempted to control the oligonucleotide surface density to investigate DNA-DNA interaction (30). A similar concern can be raised when analyzing DNA-protein

interaction in the context of molecular crowding on the surface. Thus, in order to obtain an accurate comparison of DNA-protein interaction between two different types of DNA array (ssDNA and dsDNA), the initial oligonucleotide surface densities must be comparable. Our previous work showed that the surface density of ssDNA arrays was two-fold higher than that of dsDNA arrays when both ssDNA and dsDNA were directly immobilized on copoly (DMA-NAS-MAPS)-functionalized surfaces (24). For this work, it is important to realize that the ssDNA arrays were prepared in a way that resulted in a probe density similar to that of the dsDNA arrays. The probe densities of ssDNA are all within ±30% (Fig. 3 b) of those of the corresponding dsDNA sequences. The effect of different surface densities of the DNA probes on TBP-DNA interaction is shown in Fig. S2.

Multiple TBPs bind to single-stranded poly-T

To determine the binding trend of TBPs relative to the number of repetitive Ts, we investigated TBP binding to poly-T oligonucleotides of different lengths. Four ssDNA arrays containing 6-mer, 9-mer, 12-mer, 15-mer, 18-mer, 25-mer, and 40-mer poly-T were prepared. Binding experiments were carried out by incubating the DNA arrays with 10 μg/mL of TBP for 2 h. The ratio between the ssDNA probes to the bound TBP was calculated, and the results are shown in Fig. 4. TBP did not interact with 6-mer poly-T, suggesting that TBP requires more than six nucleotides to interact. The increasing number of TBPs bound with respect to the increasing length of poly-T oligonucleotides suggests that multiple TBPs are bound along the length of the DNA strand, and the binding signal is not due to TBP aggregation.

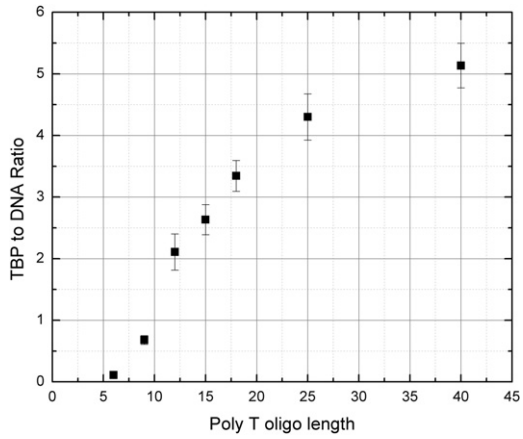


FIGURE 4 Ratio of the number of TBPs bound per poly-T sequence for different lengths. The ratios of TBP to DNA of 6-mer, 9-mer, 12-mer, 15-mer, 18-mer, 25-mer, and 40-mer poly-T (n = 9–12) were averaged per array, and the mean values across all four microarrays are plotted. The error bars represent ±1 standard error of four microarrays.

TBP binds to ss25-mer T with higher affinity than to dsTATA

We performed a competitor gel-shift assay with ss25-mer T and dsTATA (Fig. 5) to investigate the relative affinity of TBP to the poly-T stretches. Lane 2 shows the ss25-mer T-TBP complex, which migrates more slowly than the control without TBP in lane 1. Addition of cold ssTATA showed only partial competition in lane 4. However, when cold ssTATA was added to the dsTATA-TBP complex, the signal was markedly reduced, as shown in lane 9. This finding suggests that TBP has a higher affinity for ss25-mer T than for dsTATA. Additionally, successful competition of ss25-mer T-TBP with ssTATA and dsTATA implies that the TBP-ss25-mer T interaction is reversible. Comparing lanes 4 and 5, we argue that TBP has a higher affinity for dsTATA than for ssTATA. Taken together, these data strongly suggest that TBP binds to ssTATA with low affinity, to dsTATA with high affinity, and to ss25-mer T with the highest affinity. We performed a series of experiments with TBP binding to oligonucleotide arrays to compare the relative affinity with limited TBP concentrations (2.5, 1.0, 0.5, and 0.25 μg/mL). The high affinity of TBP for poly-T stretches was also observed with IRIS, and there was no TBP binding to ss25-mer A as expected (Fig. S3).

Using IRIS, we determined that ~4 TBP molecules bound to ss25-mer T. However, the gel mobility results show a shifted band corresponding to one TBP molecule. This discrepancy can be attributed to the different conditions used for the two techniques. For example, IRIS is run under equilibrium conditions, whereas EMSA is run under non-equilibrium conditions. We note that ss25-mer T has a slower migration than dsTATA. This difference in migration suggests that ss25-mer T has secondary structure that may contribute to TBP binding. Given the high-affinity binding of TBP to this probe, and the prevalence of repetitive T in the genome, it will be important to derive the secondary structure of ss25-mer T and obtain an x-ray or NMR structure of the TBP-ss25-mer T complex that can be compared with what we already know about the TBP-dsTATA box interface.

	ss25mer T					dsTATA				
Lane number	1	2	3	4	5	6	7	8	9	10
^{32}P DNA	+	+	+	+	+	+	+	+	+	+
TBP	-	+	+	+	+	-	+	+	+	+
Competitor ss25mer T	-	-	+	-	-	-	-	+	-	-
Competitor ssTATA	-	-	-	+	-	-	-	-	+	-
Competitor dsTATA	-	-	-	-	+	-	-	-	-	+

FIGURE 5 EMSAs of TBP with various oligonucleotides. The gel-competition assay demonstrates that TBP binds to ss25-mer T with higher affinity than to dsTATA.

Poly-T stretches are prevalent in the human core promoter region

Our search in 2000 bp of 17,920 promoters for poly-T stretches uncovered 6356 (35.5%) genes with one or more poly-T stretches. We also searched for a TATA-like element in the core promoter region and found 5162 (28.8%) genes. We categorized the relations between TATA box and poly-T stretches into three types (Fig. 6) to understand the positional dependence of poly-T stretches to the TATA box. Type I genes have TATA-less promoter and poly-T stretches in the core promoter regions. This suggests that a poly-T stretch could serve the function of a TATA box when the promoter region lacks a TATA box. We then searched for genes that have both poly-T stretches and a TATA element in the promoter regions. In type II genes, both elements are close together in the core promoter region. Considering the high affinity of TBP for poly-T stretches, the presence of both elements in the core promoter region may facilitate the initial TBP interaction to the promoter. In type III genes, the TATA box and poly-T stretches are farther apart. Table 2 lists the number of genes in each type. A complete list of type I–III genes can be found in the Supporting Material.

The TBP interactions with dsTATA box, ssTATA box, and poly-T stretches shown by IRIS were confirmed by EMSA. Furthermore, a competition assay revealed that TBP has a high affinity for poly-T stretches comparable to that for the TATA box. This interaction between TBP and poly-T stretches has several implications for transcription initiation. First, a poly-T stretch can serve as an alternative binding motif for TBP in TATA-less promoters. Our computational analysis revealed that 51 TATA-less promoters in human genome have poly-T stretches. Second, poly-T stretches within the core promoter with TATA box may upregulate gene expression by increasing the local TBP concentration in the promoter region. There is evidence that a certain length of repetitive thymine correlates with transcription activity. For example, Krebs et al. (31) identified a minimal core promoter for JC virus that functions sufficiently as an autonomously active initiator. This minimal core promoter

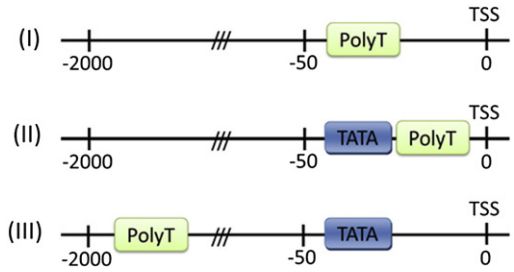


FIGURE 6 Three types of TATA-box-poly-T-stretch relation. The type I genes have TATA-less promoters and a poly-T stretches in the core promoter regions. The type II genes have a poly-T stretch with a TATA box in the core promoter region. The type III genes have a TATA box in the core promoter region and at least one poly-T stretch in the further promoter region.

TABLE 2 Number of genes in each type

	Number of genes
Type I	51
Type II	47
Type III	1865

consists of only the TATA box and an 8 bp poly-T immediately upstream. In addition, Grishkevich et al. (32) reported that the presence of T-blocks (3–5 poly-T sequence) in the core promoter positively correlated with gene expression levels in *Caenorhabditis elegans*, and that T-blocks may be associated with nucleosome eviction, clearing the promoter for transcription. The computational results suggest that the average distance between the TATA box and poly-T stretch in these 51 TATA-less promoters is ~15 bp, and 21.6% of them are within 5 bp. Finally, poly-T stretches farther away from the TATA box may have a regulatory role in TBP-initiated PIC formation. We found ~2000 genes in the human genome that contain a poly-T stretch 2000 bp upstream of TSS, which also has a TATA box in the core promoter. We propose two possible regulatory roles: 1) A long poly-T stretch could act as a sink for TBP and decrease the TPB concentration in the core promoter region, which would result in decreased TBP-dependent PIC formation and downregulation of the gene expression. 2) Poly-T stretches could facilitate TBP binding to the TATA box by having the TBP bypass the very slow 3D diffusion and delivering TBP to the promoter region through protein hopping and intersegment transfer. Regulatory proteins have been shown to undergo translocation on DNA in search of specific binding sites through macroscopic dissociation-reassociation processes (33–35).

CONCLUSIONS

We have presented a novel (to our knowledge) method for investigating TF-DNA interaction with IRIS, an interferometry-based label-free sensor with a high-throughput capacity. We demonstrated a simple method for fabricating ssDNA and dsDNA arrays with comparable probe density, enabling a correct comparison of the interactions among various types of DNA probes across multiple arrays. Maintaining a comparable probe density is particularly important in quantitative analysis of TF-DNA interactions on a surface, and we show that surface oligonucleotide density can influence the protein interaction to the surface probes. The number of probes investigated in this work was limited because we sought to demonstrate the feasibility of IRIS as a platform to study TF; however, high throughput can easily be incorporated because the throughput of IRIS is restricted only by the size of the camera (16). IRIS allowed parallel investigation of multiple DNA probe interactions with a given TF. By controlling the concentration of the protein, we determined the relative affinity of TBP for

different oligonucleotide sequences, and these results were confirmed with traditional gel electrophoresis techniques. The discovery of TBP's high affinity for poly-T stretches led to a computational genome analysis, and we found that poly-T stretches are abundant (35.5%) in promoter regions. Based on the positional analysis of poly-T stretches and TATA box, we found 51 TATA-less promoters that can be regulated by TBP through the presence of poly-T stretch in the core promoter. In addition, we suggest a regulatory role of poly-T stretches in transcription initiation for genes that present both poly-T stretch and a TATA element. Further studies, particularly in vitro investigations of poly-T stretches, will help us understand the diverse mechanisms of eukaryotic transcription initiation.

SUPPORTING MATERIAL

Additional details are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00926-5.

This work was supported by the National Institutes of Health (grant GM074872-02).

The authors thank Professor Richard Laursen, Dr. David Bergstein, Dr. Carlos Lopez, George Daaboul, Dr. Phil Spuhler, Elif Cevik, Rebecca Benham, Julia Kim, Kristen Hokenson, Meaghan Cogswell, and Zhuting Li for intellectual discussions.

REFERENCES

1. Smale, S. T., and J. T. Kadonaga. 2003. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* 72:449–479.
2. Sandelin, A., P. Carninci, …, D. A. Hume. 2007. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat. Rev. Genet.* 8:424–436.
3. Juven-Gershon, T., J. Y. Hsu, …, J. T. Kadonaga. 2008. The RNA polymerase II core promoter—the gateway to transcription. *Curr. Opin. Cell Biol.* 20:253–259.
4. Yang, C. H., E. Bolotin, …, E. Martinez. 2007. Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene.* 389:52–65.
5. Müller, F., M. A. Demény, and L. Tora. 2007. New problems in RNA polymerase II transcription initiation: matching the diversity of core promoters with a variety of promoter recognition factors. *J. Biol. Chem.* 282:14685–14689.
6. Wright, K. J., M. T. Marr, 2nd, and R. Tjian. 2006. TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc. Natl. Acad. Sci. USA.* 103:12347–12352.
7. Wiley, S. R., R. J. Kraus, and J. E. Mertz. 1992. Functional binding of the "TATA" box binding component of transcription factor TFIID to the -30 region of TATA-less promoters. *Proc. Natl. Acad. Sci. USA.* 89:5814–5818.
8. Carninci, P., A. Sandelin, …, Y. Hayashizaki. 2007. Genome-wide analysis of mammalian promoter architecture and evolution (vol 38, pg 626, 2006). *Nat. Genet.* 39:1174.
9. Kim, J. W., and V. R. Iyer. 2004. Global role of TATA box-binding protein recruitment to promoters in mediating gene expression profiles. *Mol. Cell. Biol.* 24:8104–8112.
10. Bulyk, M. L. 2007. Protein binding microarrays for the characterization of DNA-protein interactions. *Adv. Biochem. Eng. Biotechnol.* 104: 65–85.
11. Mukherjee, S., M. F. Berger, …, M. L. Bulyk. 2004. Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat. Genet.* 36:1331–1339.
12. de Wit, E., F. Greil, and B. van Steensel. 2005. Genome-wide HP1 binding in *Drosophila*: developmental plasticity and genomic targeting signals. *Genome Res.* 15:1265–1273.
13. Warren, C. L., N. C. S. Kratochvil, …, A. Z. Ansari. 2006. Defining the sequence-recognition profile of DNA-binding molecules. *Proc. Natl. Acad. Sci. USA.* 103:867–872.
14. Pirri, G., F. Damin, …, L. E. Depero. 2004. Characterization of a polymeric adsorbed coating for DNA microarray glass slides. *Anal. Chem.* 76:1352–1358.
15. Yalçin, A., F. Damin, …, M. S. Unlü. 2009. Direct observation of conformation of a polymeric coating with implications in microarray applications. *Anal. Chem.* 81:625–630.
16. Ozkumur, E., J. W. Needham, …, M. S. Unlü. 2008. Label-free and dynamic detection of biomolecular interactions for high-throughput microarray applications. *Proc. Natl. Acad. Sci. USA.* 105:7988–7992.
17. Ozkumur, E., S. Ahn, …, M. S. Unlü. 2010. Label-free microarray imaging for direct detection of DNA hybridization and single-nucleotide mismatches. *Biosens. Bioelectron.* 25:1789–1795.
18. Ozkumur, E., A. Yalçin, …, M. S. Unlü. 2009. Quantification of DNA and protein adsorption by optical phase shift. *Biosens. Bioelectron.* 25:167–172.
19. Spuhler, P. S., J. Knezevic, …, M. S. Unlü. 2010. Platform for in situ real-time measurement of protein-induced conformational changes of DNA. *Proc. Natl. Acad. Sci. USA.* 107:1397–1401.
20. Holmbeck, S. M. A., H. J. Dyson, and P. E. Wright. 1998. DNA-induced conformational changes are the basis for cooperative dimerization by the DNA binding domain of the retinoid X receptor. *J. Mol. Biol.* 284:533–539.
21. Boskovic, J., A. Rivera-Calzada, …, O. Llorca. 2003. Visualization of DNA-induced conformational changes in the DNA repair kinase DNA-PKcs. *EMBO J.* 22:5875–5882.
22. Daaboul, G. G., R. S. Vedula, …, M. S. Ünlü. 2011. LED-based interferometric reflectance imaging sensor for quantitative dynamic monitoring of biomolecular interactions. *Biosens. Bioelectron.* 26: 2221–2227.
23. Cretich, M., G. Pirri, …, M. Chiari. 2004. A new polymeric coating for protein microarrays. *Anal. Biochem.* 332:67–74.
24. Ahn, S., P. S. Spuhler, …, M. S. Ünlü. 2012. Quantification of surface etching by common buffers and implications on the accuracy of label-free biological assays. *Biosens. Bioelectron.* 36:222–229.
25. Chen, M., and A. von Mikecz. 2005. Formation of nucleoplasmic protein aggregates impairs nuclear function in response to SiO2 nanoparticles. *Exp. Cell Res.* 305:51–62.
26. BIOBASE. 2005. Available from http://www.biobase-international.com/product/transcription-factor-binding-sites. Accessed February, 2012.
27. Durbin, R. 1998. Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press, Cambridge/New York.
28. Peterson, A. W., R. J. Heaton, and R. M. Georgiadis. 2001. The effect of surface probe density on DNA hybridization. *Nucleic Acids Res.* 29:5163–5168.
29. Ricci, F., R. Y. Lai, …, J. J. Sumner. 2007. Effect of molecular crowding on the response of an electrochemical DNA sensor. *Langmuir.* 23:6827–6834.
30. Chen, S., M. F. Phillips, …, L. M. Smith. 2009. Controlling oligonucleotide surface density in light-directed DNA array fabrication. *Langmuir.* 25:6570–6575.

31. Krebs, C. J., M. T. McAvoy, and G. Kumar. 1995. The JC virus minimal core promoter is glial cell specific in vivo. *J. Virol.* 69:2434–2442.

32. Grishkevich, V., T. Hashimshony, and I. Yanai. 2011. Core promoter T-blocks correlate with gene expression levels in *C. elegans*. *Genome Res.* 21:707–717.

33. von Hippel, P. H., and O. G. Berg. 1989. Facilitated target location in biological systems. *J. Biol. Chem.* 264:675–678.

34. Winter, R. B., O. G. Berg, and P. H. von Hippel. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli* lac repressor—operator interaction: kinetic measurements and conclusions. *Biochemistry.* 20:6961–6977.

35. von Hippel, P. H. 2007. From "simple" DNA-protein interactions to the macromolecular machines of gene expression. *Annu. Rev. Biophys. Biomol. Struct.* 36:79–105.